



## Scale effect and bimodality in the frequency distribution of species occupancy

C. Hui

*Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa. Tel: +27-21-808-3413; Fax: +27-21-808-2995; E-mail: chui@sun.ac.za*

**Keywords:** Co-distribution network, Community assemblage, Macroecology, Nestedness, Rank curve, Raunkiaer's law of frequency, Species-by-site matrix.

**Abstract:** Community patterns in species-by-site matrices provide valuable clues for inferring ecological processes at work. One such pattern is the occupancy frequency distribution (OFD) depicting the frequency distribution of row sums (i.e., occupancy) with a quarter OFDs of bimodal forms. Another pattern that also reflects the structure of row sums is the ranked species occupancy curve (RSOC), and has been shown to imply a 50% of bimodality in OFDs. The use of RSOCs has been advocated in literature over the OFD based on two conclusions from a 6-model inference using only 24 matrices: (i) RSOCs have two general forms, with half representing bimodal OFDs; (ii) there are no effects of spatial and study scales on RSOCs of different forms. Using a much more representative dataset of 289 matrices, I cast doubt on these two conclusions. A missing but dominant RSOC model (the truncated power law) is added. The number of species and the nestedness of the community differ significantly among matrices of different RSOC forms; however, the number of sites and the taxa in the studies do not differ among RSOC or OFD forms. The quarter OFDs of bimodal forms is reassured, with the least frequent occupancy consistent with Raunkiaer's law of frequency. Importantly, a RSOC is mathematically transferrable to an OFD, with the derivative of the occupancy ranking curve being equal to the negative reciprocal of the occupancy frequency. Based on the type of the community (null versus interactive) and site environment (homogenous versus heterogeneous), four scenarios are needed to identify pre-inferring assemblage mechanisms. The results highlight the need for shifting research from the emphasis of marginal sums to the analysis of matrix structure for an in-depth understanding of the community assemblage patterns and mechanisms.

**Abbreviations:** AIC–Akaike Information Criterion; NODF–Nestedness metric based on Overlap and Decreasing Fill; OFD–Occupancy Frequency Distribution; PEF–Power Exponential Function; RSOC–Ranked Species Occupancy Curve.

### Introduction

Probing mechanisms and processes that give rise to community assemblage patterns is the jewel of ecology. As a standard practice, original data from community ecology are often recorded as a species-by-site matrix, with rows indicating species, columns different sites and the element the number of individuals (or presence/absence) of a species in a specific site. Such species-by-site matrices can then be used for constructing multiple macroecological and community assemblage patterns (e.g., species co-distributions; Gotelli and Graves 1996, Bell 2003) and inferring processes behind the community structure (Gaston and Blackburn 2000, Gaston et al. 2008). Specifically, patterns depicting the relationship between rows (*R*-mode analysis) are mainly used for detecting interactions and associations between species, whilst patterns of the relationship between columns (*Q*-mode analysis) are used for sorting habitat hierarchies and clusters (Legendre and Legendre 1998, Arita et al. 2008). For instance, the binary format of the matrix, indicating whether or not a species occurs in a site, has been widely used for calculating the co-occurrence patterns of species pairs which can then be compared with the patterns generated from null models for testing of the signal of competition in communities (Connor and Simberloff 1979, Gotelli 2000, Gotelli and McCabe 2002). Evidently, a single species-by-site matrix

can produce multiple interlinked community patterns, and mechanistic models proposed for explaining these patterns are often tested by comparing model predictions with the observed pattern via fitting a regression curve. However, the versatility of community patterns and the inherent weakness of testing contending models by curve fitting call for a more integrated way forward (McGill et al. 2007, Storch et al. 2008, Hui et al. 2009).

In a binary species-by-site matrix, the pattern depicting the frequency distribution of row sums (defined as the occupancy of a species) has been coined the occupancy frequency distribution (OFD) and is equivalent to the species-range size distribution at regional scales (Gaston 1996, Collins and Glenn 1997, McGeoch and Gaston 2002). Raunkiaer's (1934) law of frequency portrays a bimodal OFD of plant communities, suggesting species in a community are either rare or common, with only few species having intermediate occupancies. Specifically, when the occupancy is divided into five classes (0–20%, 21–40%, 41–60%, 61–80% and 81–100%), the two modes (a mode is defined as the class that has more species than its adjacent classes) appear in the first and last classes, with the valley (the least frequent occupancy) representing the class of 61–80% of all sites (McIntosh 1962). Evidently, this J-shaped OFD implies that the gradient (or the derivative) of the frequency distribution

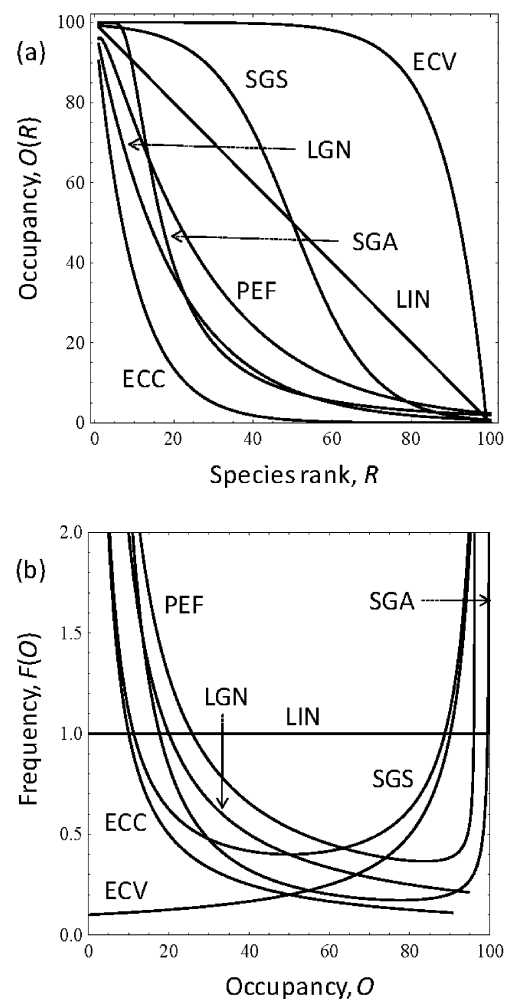
equals zero at the valley. Although the form of OFD can be diverse, the bimodality is still common (27% of the 48 matrices reviewed by McGeoch and Gaston (2002)).

Three explanations of the bimodality in OFDs include: first, an artefact from sampling highly skewed relative abundance distribution (Nee et al. 1991, Papp and Izsák 1997). For instance, Papp and Izsák (1997) demonstrate that repeated sampling with replacement from a community with either a truncated lognormal or log-series relative abundance distribution will lead to a bimodal OFD, echoing Preston's (1948) and Williams' (1950) concordance that the number of individuals per species in a community is the crux of the shape of OFDs. This line of research becomes especially relevant now since the empirical relative abundance distributions in large assemblages often have more rare species than predicted from lognormal models, indicating a mixture of species of two or multiple functions (Magurran and Henderson 2003, Hui et al. 2011a). Second, the core-satellite hypothesis states that if local extinction is subject to a strong rescue effect in a meta-community, the occupancy dynamics that is balanced by the colonization of new species and the extinction of existing species will lead to a bimodal OFD (Hanski 1982). However, the valley as predicted from the core-satellite model (eqn. 7 in Hanski 1982) represents an occupancy less than 50% of all sites, contradicting to the typical J-shaped bimodal OFD. Finally, the bimodality in OFDs could represent a transient pattern due to the effect of spatial scales on species occupancy (Hui and McGeoch 2007a, b, 2008), echoing Gleason's (1929) criticism that Raunkiaer's law of frequency only applies at certain scales.

Recently, Jenkins (2011) advocates a simple R-mode pattern (a pattern describing the row sums), namely the 'ranked species occupancy curves (RSOCs)', over the OFD in conjecturing processes giving rise to the community structure reflected in a binary species-by-site matrix. Based on the selection of the best among six RSOC models using only 24 matrices, Jenkins (2011) reaches two main conclusions. First, RSOCs only have two general forms (half exponential concave and half sigmoidal), suggesting that 50% of OFDs could be bimodal (a much higher estimate than the 27% reviewed by McGeoch and Gaston (2002)). Second, spatial and study scales of the matrices do not differ significantly among different RSOC forms, contrasting the strong effect of scales on the forms of OFDs (Hui and McGeoch 2007a, b). Here, using a much more representative dataset of 289 species-by-site matrices, I show that these two conclusions could be biased. By further adding an important RSOC model which emerged from recently literature on ecological networks, Jenkins' conclusions are modified drastically. Based on these results and further because R-mode patterns (both RSOCs and OFDs) are insensitive to species changing sites while keeping the occupancy constant, I envisage a need for shifting research focus from the emphasis on marginal sums of columns and rows (e.g., Q- and R-mode analysis) to the analysis of matrix structure (e.g., nestedness and compartmentalization) in this species-by-site co-distribution network.

## Forms of ranked species occupancy curves

For assessing the proportion of different forms of RSOCs in nature, here I used the dataset comprising 289 species-by-site matrices compiled by Atmar and Patterson (1995) from published literature, ranging from 3 to 202 sites and 3 to 309 species. Since species-by-site matrices have been used for recording sampling results from multiple sites for more than a century in ecology, the quality of these matrices, as published, is unquestionable and will not be discussed here for convenience. A RSOC can be constructed by sorting the occupancies of species from high to low and then plotting species occupancy as a function of its ranking (Fig. 1a). Following Jenkins' (2011) procedure exactly, I used the Levenberg-Marquardt algorithm of nonlinear regression (< 999 iterations) to fit six RSOC models (Fig. 1a):



**Figure 1.** An illustration of the seven models for (a) ranked species occupancy curves and (b) occupancy frequency distributions. See Table 1 for acronyms. The ranked species occupancy curves include ECC:  $O = 0.01 + 100e^{-0.1R}$ ; ECV:  $O = 100 + 0.1(1 - 0.05e^{0.1R})$ ; LGN:  $O = e^{-0.05R + 4.6}$ ; SGS:  $O = 100/(1 + e^{0.1R - 5})$ ; SGA:  $O = 100(1 - e^{-200R})$ ; LIN:  $O = 100 - R$ ; PEF:  $O = 100R^{0.05}e^{-0.04R}$ . The occupancy frequency distributions are derived from the ranked species occupancy using eqn. (2).

Exponential concave, ECC:  $O(R) = y_0 + ae^{-bR}$ ,

Exponential convex, ECV:  $O(R) = O(1) + y_0(1 - ae^{bR})$ ,

Lognormal function, LGN:  $O(R) = \exp(aR + b)$ ,

Sigmoidal symmetric, SGS:  $O(R) = a/(1 + e^{-bR+c})$ ,

Sigmoidal asymmetric, SGA:  $O(R) = a/(1 + e^{-bR^c})$ ,

Linear, LIN:  $O(R) = aR + b$ ,

where  $O(R)$  is the occupancy of the species ranked  $R$  in the community, and  $y_0$ ,  $a$ ,  $b$  and  $c$  are coefficients. These models are not originally proposed for depicting RSOCs *per se*. For instance, the two sigmoidal models (SGS and SGA) were initially proposed for depicting species-area curves (Tjørve 2003). The lognormal function was proposed for the rank-abundance curve (Loehle and Hansen 2005). However, these models represent a variety of possible forms of monotone functions and are, thus, suitable for depicting RSOCs.

Because these models have different numbers of parameters, Akaike information criterion (AIC) was used to select the best model that fits the empirical data (Burnham and Anderson 2002). The model with the highest Akaike information criterion (AIC) weight was assigned as the form of RSOC to a specific species-by-site matrix. The results exhibited a systematic deviation from Jenkins' results (Table 1; see also Appendix S1 in the Online Attachment for detailed results). Jenkins (2011) identifies three models for the RSOCs of 24 matrices, contrasting five models identified for the 289 matrices when using the same six models (Jenkins vs. 6-model in Table 1), with less SGA, no SGS but higher proportions for other 4 models. Importantly, this means much less sigmoidal but more lognormal and linear forms of RSOCs in these 289 matrices (Table 1).

As to the nature of the AIC analysis, the model selected represents only the best among contending models; adding a new contending model will strongly affect the results. Although this does not advocate for testing any possible models, it is essential to include those with clear underpinnings of ecological processes and substantial empirical support in the multi-model inference. One important RSOC model missing in Jenkins' (2011) analysis is the power exponential function, PEF (Fig. 1a; also known as the truncated power law):

$$O(R) = aR^b e^{-cR}. \quad (1)$$

The PEF model consists of two parts: a power-law function ( $aR^b$ ) and an exponential cut-off ( $e^{-cR}$ ). The power-law function indicates the scale-free structure that the relationship between species ranks and their occupancies does not stand out at a particular scale (Schroeder 1991), representing a heterogeneous (fractal) habitat. The exponential cut-off indicates a Poisson process that species occupancy is not site-specific and thus represents a homogeneous habitat. For widespread species (i.e., when  $R$  is small), the PEF model is largely controlled by the power-law function (that is, widespread species perceive the sites differentially); for rare species (i.e., when  $R$  is large), the shape is dominated by the exponential cut-off

**Table 1.** Proportions of models for ranked species occupancy curves (ECC, exponential concave; ECV, exponential convex; LGN: lognormal; SGS: sigmoidal symmetric; SGA: sigmoidal asymmetric; LIN: linear; PEF: power exponential function) and proportions of bimodality in occupancy frequency distributions. 'Jenkins' and '6-model' indicate results without the PEF model for the 24 matrices in Jenkins (2011) and the 289 matrices analyzed here. '7-model' indicates results with the PEF model added for the 289 matrices.

Model	Jenkins	6-model	7-model
ECC	0.46	0.61	0.35
ECV	-	0.04	0.03
LGN	-	0.20	0.11
SGS	0.25	-	-
SGA	0.29	0.03	0.01
LIN	-	0.12	0.07
PEF			0.44
Bimodality	0.54	0.03	0.24

(that is, rare species occur in a site by chance). The PEF model has also been widely observed when depicting emerging patterns in many ecological networks. For instance, interactions in a plant-animal mutualistic network (e.g., flowering plants and their pollinators) can be depicted by a binary connectivity matrix, and the distribution of node degrees (equals the occupancy here) predominately follows a PEF ranking form (e.g., Bascompte and Jordano 2007).

Adding this PEF model into the 6-model selection drastically altered the proportion of each RSOC model in these 289 matrices (6- vs. 7-model in Table 1). The PEF model had the highest proportion, taking half the share from the ECC model; the linear and lognormal models remained substantial (Table 1). Overall, sigmoidal models (SGS and SGA) became neglectable; the PEF model turns out most common; linear and lognormal models are not rare (Table 1). Reasons for this vast discrepancy are potentially due to that (i) Jenkins' results are not statistically representative, with the potential of overfitting 6 models in only 24 matrices, and (ii) missing out an important contending model.

### From sigmoidal form to bimodality

Here, I first show the mathematical link between a RSOC and its corresponding OFD. Let  $O(R)$  be the occupancy of a species with a ranking of  $R$ ,  $R(O)$  the ranking of a species with occupancy  $O$ , and  $F(O)$  the number of species (i.e., the frequency) of occupancy  $O$ . Evidently, the rank of a species is equal to one plus the number of species with greater occupancies than itself,

$$R(O) = 1 + \int_O^N F(x) dx,$$

where  $N$  is the total number of sites (i.e., the maximum occu-

pancy). Therefore, we have the derivative of the species' rank with respect to its occupancy,  $R'(O) = -F(O)$ . Since  $R(O)$  is the inverse function of  $O(R)$ , we have  $R'(O) = 1/O'(R)$ , and thus the derivative of the occupancy ranking curve can be expressed as the negative reciprocal of the occupancy frequency,

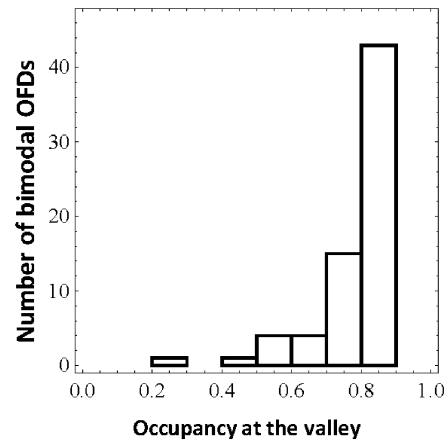
$$O'(R) = -1/F(O(R)) \quad (2)$$

This means that the RSOC and the OFD are mathematically transferable; preference of one over the other seems to be groundless. The above equation provides a general relationship between rank curves and frequency distributions (Fig. 1b). A parallel application is to infer species abundance distributions from their corresponding rank abundance curves, and vice versa. For instance, Izsák and Pavoine (2012) have demonstrated a connection between the forms of species abundance distributions and the concavity and convexity of their rank abundance curves. We can further have  $O''(R) = -F(O)/F(O(R))^3$ , indicating that a bimodal OFD ( $F'(\hat{O}) = 0$  at the occupancy valley  $\hat{O}$ ) is mathematically equal to a sigmoidal RSOC (i.e., the existence of an inflection point  $O'' = (\hat{R}) = 0$ ). This then further allows us to detect bimodal OFDs in these species-by-site matrices.

Among the six models in Jenkins (2011), only the SGS and SGA models have inflection points ( $\hat{O} = a/2$  and  $a(1-e^{-(1-c)/c})$ , respectively) and thus represent bimodal OFD. Jenkins (2011) suggests that 13 out of 24 matrices (54%) exhibit sigmoidal RSOC and thus bimodal OFD. In contrast, when the six models were examined for the 289 matrices, only 5 SGA models were identified, suggesting a mere 3% of bimodal OFDs (Table 1). By adding the new PEF model into the multi-model inference, nearly half of the ECC forms changed to the PEF form. Because the PEF model can potentially have an inflection point ( $\hat{O} = a((\sqrt{b} + b)/c)^b e^{-\sqrt{b}-b}$ , if  $b > 0$ ), 68 of the 127 PEF forms of RSOCs were actually bimodal. This yielded an estimate of 24% bimodality in OFDs, consistent with McGeoch and Gaston's (2002) estimate of 27% bimodality. It is clear that the un-representativeness of Jenkins' 24 matrices and the missing of the PEF model have led to an estimate of 50% bimodality in OFDs. One quarter matrices showing a bimodal OFD should represent the reality.

### Effect of scales on occupancy ranking forms

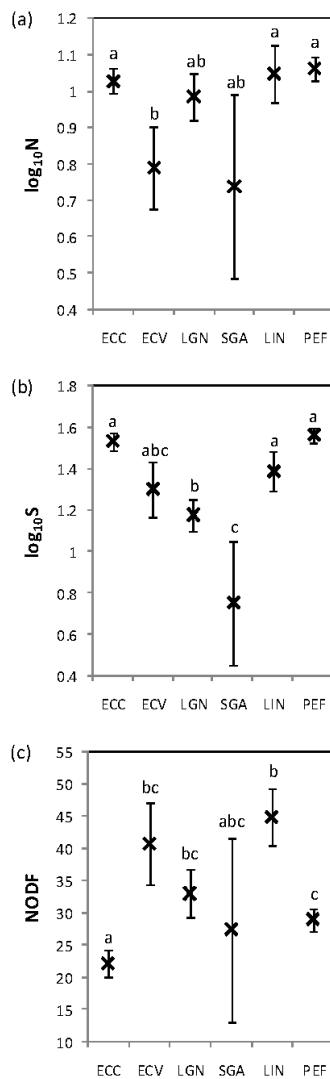
Jenkins (2011) concludes that RSOC is a better community pattern than OFD because communities of different RSOC forms do not differ significantly in their spatial (grain and extent) and study scales (number of sites and species). This is again misleading. Species occupancy and thus the forms of OFDs are highly scale dependent (Gleason 1929, Gaston 1996, McGeoch and Gaston 2002, Hui et al. 2006, 2010, 2011b), so should be the RSOC forms given that the OFD and RSOC are mathematically transferable. Failure to detect the scale effect could be due to the un-representativeness of the 24 matrices. When examined using the 289 matrices, most occupancies at the valley for the 68 bimodal-OFD matrices ranged between 0.8 and 0.9, with the rest



**Figure 2.** The frequency distribution of the occupancy at the valley (i.e., the least frequent occupancy) for the 68 bimodal occupancy frequency distributions.

mainly between 0.5 to 0.8 and none above 0.9 (Fig. 2), consistent with both Raunkiaer's law of frequency and the prediction from a model depicting the scale effect on species occupancy (see Fig. 5g-i in Hui and McGeoch 2007a), suggesting that a bimodal OFD could reflect highly aggregated (fractal) distributions of species but with trivial inter-specific interactions in the community.

Using ANOVA, I found that both the 6-model and 7-model comparison yielded consistent results: the log-transformed number of sites ( $\log_{10}N$ ) do not differ significantly among different RSOC models (6-model:  $F_{4,284} = 1.74$ ,  $P = 0.14$ ; 7-model:  $F_{5,283} = 1.49$ ,  $P = 0.19$ ), but the log-transformed number of species ( $\log_{10}S$ ) differed significantly (6-model:  $F_{4,284} = 9.75$ ,  $P < 0.01$ ; 7-model:  $F_{5,283} = 6.26$ ,  $P < 0.01$ ). Since the species-by-site matrix is in the same format as the incidence matrices in ecological networks, it is reasonable to further check how the nestedness of these matrices corresponds to different RSOC forms. A highly nested matrix, often measured by the Nestedness metric based on Overlap and Decreasing Fill (NODF; Almeida-Neto et al. 2008), indicates that a species only occupies a subset of where a species with a higher occupancy occurs but does not occur in a different combination of sites, forming a nested structure of co-distribution. Interestingly, the NODF of nestedness (Guimarães and Guimarães, 2006) also differed significantly among different models of RSOCs (6-model:  $F_{4,284} = 4.72$ ,  $P < 0.01$ ; 7-model:  $F_{5,283} = 5.77$ ,  $P < 0.01$ ). A pairwise *t*-test further revealed that the convex RSOCs normally had a lower number of sites but a higher nestedness than concave RSOCs, whereas asymmetric sigmoidal RSOCs had a significantly lower number of species than others (Fig. 3). Following Jenkins (2011), I found that the taxonomic groups with more than 10 matrices (birds, fishes, mammals, reptiles, plants, snails and arthropods) showed no significant differences from expectation ( $\chi^2$  test,  $P > 0.19$ ), except when ants were excluded in arthropods ( $\chi^2 = 12.3$ ,  $df = 5$ ,  $P = 0.03$ ).



**Figure 3.** Pairwise comparison of log-transformed number of sites (a), species (b), and the nestedness (c) measured by NODF, for matrices with different forms of ranked species occupancy curves. See Table 1 for acronyms; letters are significant differences ( $t$ -test;  $P < 0.05$ ), with ANOVA already confirming significant differences in (b) and (c).

## Discussion

Both RSOCs and OFDs are marginal patterns of the species-by-site matrices; therefore, they can neither distinguish the effect of site heterogeneity and hierarchies, nor species' niche (site) preference and interactions. It is difficult to pinpoint one of the following four scenarios to a community by simply examining the forms of RSOCs and OFDs. First, if there are no site preference (i.e., the occurrence of a species in a specific site depends solely by chance) and no species interaction (i.e., the occurrence of one species is not associated with the occurrence of other species), we can expect a same number of species per site and a null community (i.e., species independence) in a homogeneous landscape. Second, without site preference but with species interaction, species

in this interactive community can form hierarchies (a nested community) or functional modules (a compartmentalised community) even in a homogeneous landscape. Third, with site preference but without species interaction, a null community in a heterogeneous landscape will emerge (specifically, the same site/niche preference leads to a neutral community). Finally, with both site preference and species interaction, we have an interactive community in a heterogeneous landscape. Once a scenario is identified for a community (via null model tests; Gotelli and Graves 1996), the commonness and rarity of a species (and thus the forms of RSOC and OFD) can then be explained as a consequences of abundance sampling (Papp and Izsak 1997), core/satellite processes (Hanski 1998), scale artefact (Hui and McGeoch 2007a, b), and/or community succession (Jenkins 2011).

Indeed, unlike bipartite networks (e.g., a pollination network with rows representing flowering plants and columns their pollinators) where the matrix structure is formed by interactions between columns and rows (e.g., Zhang et al. 2011), species-by-site matrices only reflect preference and interactions between rows (species) and thus depict a unilateral interaction (from rows to columns). This explains why the number of sites does not differ among RSOC models but the number of species does. Although Jenkins' (2011) conclusions are biased due to the low number of matrices and the missing PEF model, he does however expose the strengths and weaknesses of R-mode analysis. Borrowing Thomas Kuhn's phrase, I see a need for a drastic paradigm shift in analyzing species-by-site matrices, from the emphasis on marginal sums (e.g., R- and Q-mode analysis) to the analysis of matrix structure (e.g., nestedness and compartmentalisation). In conclusion, the effects of spatial and study scales are not trivial but evident, and the bimodality could account for a quarter of the species-by-site matrices.

**Acknowledgments.** I am grateful to C. Ricotta, B.D. Patterson, M.A. McGeoch and two anonymous referees for discussions and constructive comments on relevant issues. This work is supported by the NRF Blue Sky and Incentive Research Programme.

## References

- Almeida-Neto, M., P. Guimarães, P.R. Guimarães Jr, R.D. Loyola, and W. Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117: 1227-1239.
- Arita, H.T., J.A. Christen, P. Rodriguez, and J. Soberon. 2008. Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. *Am. Nat.* 172: 519-532.
- Atmar, W. and B.D. Patterson. 1995. *The Nestedness Temperature Calculator: A Visual Basic Program, Including 294 Presence-absence Matrices*. AICS Research, Inc., University Park, NM and The Field Museum, Chicago, IL.
- Bascompte, J. and P. Jordano. 2007. Plant-animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 38: 567-593.

- Bell, G. 2003. The interpretation of biological surveys. *Proc. R. Soc. B* 270: 2531-2542.
- Burnham, K.P. and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-theoretical Approach*. 2<sup>nd</sup> edition. Springer-Verlag, Berlin.
- Collins, S.L. and S.M. Glenn. 1997. Effects of organismal and distance scaling on analysis of species distribution and abundance. *Ecol. Appl.* 7: 543-551.
- Connor, E.F. and D. Simberloff. 1979. The assembly of species communities: chance or competition? *Ecology* 60: 1132-1140.
- Gaston, K.J. 1996. Species-range size distributions: patterns, mechanisms and implications. *Trends Ecol. Evol.* 11: 197-201.
- Gaston, K.J. and T.M. Blackburn. 2000. *Pattern and process in macroecology*. Blackwell Science, Oxford.
- Gaston, K.J., S.L. Chown, and K.L. Evans. 2008. Ecogeographical rules: elements of a synthesis. *J. Biogeogr.* 35: 483-500.
- Gleason, H.A. 1929. The significance of Raunkiaer's law of frequency. *Ecology* 10: 406-408.
- Gotelli, N.J. 2000. Null model analysis of species co-occurrence patterns. *Ecology* 81: 2606-2621.
- Gotelli, N.J. and G.R. Graves. 1996. *Null Models in Ecology*. Smithsonian Institution Press, Washington, DC.
- Gotelli, N.J. and D.J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* 83: 2091-2096.
- Guimarães, P.R. Jr and P. Guimarães. 2006. Improving the analyses of nestedness for large sets of matrices. *Environ. Model. Software* 21: 1512-1513.
- Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38: 210-221.
- Hui, C. and M.A. McGeoch. 2007a. A self-similarity model for occupancy frequency distribution. *Theor. Popul. Biol.* 71: 61-70.
- Hui, C. and M.A. McGeoch. 2007b. Modeling species distributions by breaking the assumption of self-similarity. *Oikos* 116: 2097-2107.
- Hui, C. and M.A. McGeoch. 2008. Does the self-similar species distribution model lead to unrealistic predictions? *Ecology* 89: 2946-2952.
- Hui, C., M.A. McGeoch, and M. Warren. 2006. A spatially explicit approach to estimating species occupancy and spatial correlation. *J. Anim. Ecol.* 75: 140-147.
- Hui, C., M.A. McGeoch, B. Reyers, P.C. le Roux, M. Greve, and S.L. Chown. 2009. Extrapolating population size from the occupancy-abundance relationship and the scaling pattern of occupancy. *Ecol. Appl.* 19: 2038-2048.
- Hui, C., R. Veldtman, and M.A. McGeoch. 2010. Measures, perceptions and scaling patterns of aggregated species distributions. *Ecography* 33: 95-102.
- Hui, C., L.C. Foxcroft, D.M. Richardson, and S. MacFadyen. 2011a. Defining optimal sampling effort for large-scale monitoring of invasive alien plants: a Bayesian method for estimating abundance and distribution. *J. Appl. Ecol.* 48: 768-776.
- Hui, C., D.M. Richardson, M.P. Robertson, J.R.U. Wilson, and C.J. Yates. 2011b. Macroecology meets invasion ecology: linking the native distributions of Australian acacias to invasiveness. *Diversity Distrib.* 17: 872-883.
- Izsák, J. and S. Pavoine. 2012. Links between the species abundance distribution and the shape of the corresponding rank abundance curve. *Ecol. Indic.* 14: 1-6.
- Jenkins, D.G. 2011. Ranked species occupancy curves reveal common patterns among diverse metacommunities. *Global Ecol. Biogeogr.* 20: 486-497.
- Legendre, P. and L. Legendre. 1998. *Numerical Ecology*, 2<sup>nd</sup> edition. Elsevier, Amsterdam.
- Loehle, C. and A. Hansen. 2005. Community structure and scaling relations for the avifauna of the US pacific and inland northwest. *Ecol. Complexity* 2: 59-70.
- Magurran, A.E. and P.A. Henderson. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature* 422: 714-716.
- McGeoch, M.A. and K.J. Gaston. 2002. Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biol. Rev.* 77: 311-331.
- McGill, B.J., R.S. Etienne, J.S. Gray, D. Alonso, M.J. Anderson, H.K. Benecha, M. Dornelas, B.J. Enquist, J.L. Green, F. He, A.H. Hurlbert, A.E. Magurran, P.A. Marquet, B.A. Maurer, A. Ostling, C.U. Soykan, K.I. Ugland, and E.P. White. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10: 995-1015.
- McIntosh, R.P. 1962. Raunkiaer's "law of frequency". *Ecology* 43: 533-535.
- Nee, S., R.D. Gregory, and R.M. May. 1991. Core and satellite species: theory and artefacts. *Oikos* 62: 83-87.
- Papp, L. and J. Izsák. 1997. Bimodality in occurrence classes: a direct consequence of lognormal or logarithmic series distribution of abundances – a numerical experimentation. *Oikos* 79: 191-194.
- Preston, F.W. 1948. The commonness and rarity of species. *Ecology* 29: 254-283.
- Raunkiaer, C. 1934. *The Life Forms of Plants and Statistical Plant Geography Being the Collected Papers of C. Raunkiaer*. Clarendon Press, Oxford.
- Schroeder, M. 1991. *Fractals, Chaos, and Power Laws*. Freeman, New York.
- Storch, D., A.L. Šizling, J. Reif, J. Polechová, E. Šizlingová, and K.J. Gaston. 2008. The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. *Ecol. Lett.* 11: 771-784.
- Tjørve, E. 2003. Shapes and functions of species-area curves: a review of possible models. *J. Biogeogr.* 30: 827-835.
- Williams, C.B. 1950. The application of the logarithmic series to the frequency of occurrence of plant species in quadrats. *J. Ecol.* 38: 107-138.
- Zhang, F., C. Hui, and J.S. Terblanche. 2011. An interaction switch predicts the nested architecture of mutualistic networks. *Ecol. Lett.* 14: 797-803.

Received October 10, 2011  
 Revised November 15, 2011  
 Accepted November 25, 2011

## Appendix S1

Detailed results for the 289 species-by-site matrices. The file may be downloaded from the web site of the publisher at [www.akademai.com](http://www.akademai.com).